

Deterministic replica-exchange method without pseudo random numbers for simulations of complex systems

Ryo Urano¹ and Yuko Okamoto^{1,2,3,4}

¹*Department of Physics, Graduate School of Science,
Nagoya University, Nagoya, Aichi 464-8602, Japan*

²*Structural Biology Research Center, Graduate School of Science,
Nagoya University, Nagoya, Aichi 464-8602, Japan*

³*Center for Computational Science, Graduate School of Engineering,
Nagoya University, Nagoya, Aichi 464-8603, Japan and*

⁴*Information Technology Center, Nagoya University, Nagoya, Aichi 464-8601, Japan*

We propose a replica-exchange method (REM) which does not use pseudo random numbers. For this purpose, we first give a conditional probability for Gibbs sampling replica-exchange method (GSREM) based on the heat bath method. In GSREM, replica exchange is performed by conditional probability based on the weight of states using pseudo random numbers. From the conditional probability, we propose a new method called deterministic replica-exchange method (DETREM) that produces thermal equilibrium distribution based on a differential equation instead of using pseudo random numbers. This method satisfies the detailed balance condition using a conditional probability of Gibbs heat bath method and thus results can reproduce the Boltzmann distribution within the condition of the probability. We confirmed that the equivalent results were obtained by REM and DETREM with two-dimensional Ising model. DETREM can avoid problems of choice of seeds in pseudo random numbers for parallel computing of REM and gives analytic method for REM using a differential equation.

Keywords: generalized-ensemble algorithm, replica-exchange method (REM), simulated tempering (ST), Monte Carlo (MC) simulation, differential equation, gibbs sampling, heat-bath method, conditional probability, pseudo random numbers, Ising model

INTRODUCTION

The enhancement of sampling during Monte Carlo (MC) and molecular dynamics (MD) simulations is very important for complex systems. Replica-exchange method (REM) (or parallel tempering) is one of the most popular ways to improve sampling efficiency[1–4] including biomolecular system in explicit solvent[5, 6] or biomembrane[7, 8] (for reviews, see, e.g., Refs.[9, 10]). To realize a thermal equilibrium distribution, REM uses Metropolis criterion with pseudo random numbers. However, random numbers sometimes give inaccurate results of simulations[11]. Moreover, generation of high quality random numbers is often difficult and does not assure good simulation results[12]. REM and its extension is suited for parallel computing[13–16]. Most of pseudo random number generators decrease the scalability in parallelization[17]. Hence, the complementary method producing the same results without pseudo random numbers is meaningful.

In addition, the analytic approach for temperature selections in REM have been performed [18, 19]. For performance and the condition of REM, several works were also performed. For example, Nymeyer [20] showed how efficient REM is than conventional simulations using the number of independent configurations. Abraham and Gready introduced some measurement and compared the results[21]. Rosta and Hummer [22] evaluated the practical efficiency of REM simulation for protein folding with a two-state model. However, the examination of the condition for convergence of REM is difficult partly because the mixing of temperature in REM is determined by pseudo random numbers with Metropolis criteria. As a result, most of analyses estimated the REM performance by simulation results.

Recently, Suzuki *et al.* proposed a method to produce a thermal equilibrium state without using random numbers for spin models by a differential equation based on the conditional probability of Gibbs sampling heat bath method, which is referred to as chaotic Boltzmann machines[23, 24]. The differential equation controls spin states at each site and the staying time of each spin state is proportional to the weights of the thermal equilibrium distribution. They reproduced the results of a conventional MC method in some spin systems.

Moreover, Boltzmann machine [25] has its mathematical framework [26]. The method was analyzed by mean field approximation[27], algebraic geometry and informative geometry. For example, a linear convergence of parameters in Boltzmann machine was suggested by a learning algorithm of Fisher information matrices[28], and an upper boundary for performance was obtained by algebraic geometry [29]. By introducing the differential equation for replica-exchange method, the previous results in the fields can be applied for the REM analysis. This means that analytic approach for Boltzmann machine will be applied for REM by this extension. Moreover, this new implementation of REM

will be related to hierarchical structure of Boltzmann machine, which is similar to deep Boltzmann machine [30, 31]. Developments in Boltzmann machine to accelerate convergence of sampling such as Contrastive Divergence method[32] have been proposed.

We here generalize this Chaotic Boltzmann machine to REM. We first have to extend the conditional probability for replica exchange not based on Metropolis criterion but on a Gibbs sampling heat bath method. The heat bath formalism has already been given in Ref. [33], we refer to this method as Gibbs sampling replica-exchange method (GSREM). (A similar approach based on global balance condition[34] was also developed[35].) We then introduce a differential equation for replica exchange to modify GSREM. This method is referred to as the deterministic replica-exchange method (DETREM). We then tested the effectiveness of DETREM by comparing the results of simulation of 2-dimensional Ising model with those by the conventional REM.

The organization of this paper is as follows. In Section 2, the theory for the new method and conventional REM is presented. In Section 3, we give the results of DETREM together with REM. The final section is devoted to conclusions.

METHODS

We first briefly review the conventional REM. We prepare M non-interacting replicas at M different temperatures. Let the label i ($=1, \dots, M$) stand for the replica index and label m ($=1, \dots, M$) for the temperature index. Here, i and m are related by the permutation functions by

$$\begin{cases} i = i(m) \equiv f(m), \\ m = m(i) \equiv f^{-1}(i), \end{cases} \quad (1)$$

where $f(m)$ is a permutation function of m and $f^{-1}(i)$ is the inverse. We represent the state of the entire system of M replicas by $X = \{x_{m(1)}^{[1]}, \dots, x_{m(M)}^{[M]}\}$, where $x_m^{[i]} = \{q^{[i]}, p^{[i]}\}_m$ are the set of coordinates $q^{[i]}$ and momenta $p^{[i]}$ of particles in replica i (at temperature T_m). The probability weight factor for state X is given by a product of Boltzmann factors:

$$W_{\text{REM}}(X) = \prod_{i=1}^M \exp[-\beta_{m(i)} H(q^{[i]}, p^{[i]})], \quad (2)$$

where $\beta_m (= 1/k_B T_m)$ is the inverse temperature and $H(q, p)$ is the Hamiltonian of the system. We consider exchanging a pair of replicas i and j corresponding to temperatures T_m and T_n , respectively:

$$X = \{\dots, x_m^{[i]}, \dots, x_n^{[j]}, \dots\} \rightarrow X' = \{\dots, x_m^{[j]'}, \dots, x_n^{[i]'}, \dots\}, \quad (3)$$

where $x_n^{[i]'} \equiv \{q^{[i]}, p^{[i]'}\}_n$, $x_m^{[j]'} \equiv \{q^{[j]}, p^{[j]'}\}_m$, and $p^{[j]'} = \sqrt{\frac{T_m}{T_n}} p^{[j]}$, $p^{[i]'} = \sqrt{\frac{T_n}{T_m}} p^{[i]}$ [4]. The exchange of replicas introduces a new permutation function f' :

$$\begin{cases} i = f(m) \rightarrow j = f'(m), \\ j = f(n) \rightarrow i = f'(n). \end{cases} \quad (4)$$

We remark that this process is equivalent to exchanging a pair of temperatures T_m and T_n for the corresponding replicas i and j .

Here, the transition probability $\omega(X \rightarrow X')$ of Metropolis criterion is given by

$$\omega(X \rightarrow X') = \min\left(1, \frac{W_{\text{REM}}(X')}{W_{\text{REM}}(X)}\right) = \min(1, \exp(-\Delta)), \quad (5)$$

where

$$\Delta = \Delta_{m,n} = (\beta_n - \beta_m)(E(q^{[i]}) - E(q^{[j]})). \quad (6)$$

REM is performed by repeating the following two steps:

1. We perform a conventional MD or MC simulation of replica i ($= 1, \dots, M$) at temperature T_m ($m = 1, \dots, M$) simultaneously and independently for short steps.

2. Selected pairs of replicas are exchanged based on the above Metropolis criterion in Eqs. (5) and (6). A pseudo random number is used to judge the criterion.

Without loss of generality we can assume $T_1 < T_2 < \dots < T_M$. Note that in Step 2 we usually exchange only pairs of replicas corresponding to neighboring temperatures, because the acceptance probability for replica exchange decreases exponentially with the difference of the two inverse temperatures and potential energy terms because of Eq. (6). This replica exchange can be written as

$$X = \{\dots, x_m^{[i]}, \dots, x_{m+1}^{[j]}, \dots\} \rightarrow X' = \{\dots, x_m^{[j]'}, \dots, x_{m+1}^{[i]'}, \dots\}, \quad (7)$$

where in Eq. (5) Δ is now given by

$$\Delta_m = (\beta_{m+1} - \beta_m)(E(q^{[i]}) - E(q^{[j]})). \quad (8)$$

The REM method makes a random walk in temperature space during the simulation. The canonical ensemble is reconstructed by the multiple-histogram reweighting technique, or weighted histogram analysis method (WHAM)[36, 37].

We next present GSREM[33]. As in the conventional REM, we usually consider the neighboring temperature exchange in Eq. (7). The conditional probability $\omega(x_m^{[j]}, x_{m+1}^{[i']} | x_{m(k)}^{[k \neq i(m), j(m+1)]})$, in which the new state selects the temperature exchanged state of replicas i and j with T_{m+1} and T_m from the no-exchange state of replicas i and j with temperatures T_m and T_{m+1} , is given by

$$\omega(x_m^{[j]}, x_{m+1}^{[i']} | x_{m(k)}^{[k \neq i(m), j(m+1)]}) = \frac{W(x_m^{[j]}, x_{m+1}^{[i']} | x_{m(k)}^{[k \neq i(m), j(m+1)]})}{W(x_m^{[i]}, x_{m+1}^{[j]} | x_{m(k)}^{[k \neq i(m), j(m+1)]}) + W(x_m^{[j]}, x_{m+1}^{[i']} | x_{m(k)}^{[k \neq i(m), j(m+1)]})} \quad (9)$$

$$= \frac{1}{1 + \frac{W(x_m^{[i]}, x_{m+1}^{[j]} | x_{m(k)}^{[k \neq i(m), j(m+1)]})}{W(x_m^{[j]}, x_{m+1}^{[i']} | x_{m(k)}^{[k \neq i(m), j(m+1)]})}}. \quad (10)$$

In GSREM, the above procedure for the conventional REM is performed, where Step 2 for the GSREM is performed based on Eq. (25). Here, in Step 2, the conditional probability of a temperature set based on Eq. (25) is calculated, and this assigns weights between 0 and 1 for exchanged states and a no-exchange state. Finally, after a pseudo random number is generated, the state corresponding to the random number with the assigned region is selected. For the Boltzmann distribution, this equation in Eq. (9) can be rewritten as

$$\omega(x_m^{[j]}, x_{m+1}^{[i']} | x_{m(k)}^{[k \neq i(m), j(m+1)]}) = \frac{1}{1 + \exp(\Delta_m)}, \quad (11)$$

where Δ_m is given by Eq. (8). This is the Gibbs sampling replica-exchange method when an equilibrium state is produced by this conditional probability with pseudo random numbers. We remark that REM and GSREM are mathematically equivalent in the present case because updates of new states have only two possibilities. A more general formulation for GSREM is given in the Appendix, which is mathematically different from REM. We next propose DETREM. At first, as in the conventional REM, we can only use the internal states $y_m \in \{-1, 1\}$ for a pair of neighboring temperatures (T_m, T_{m+1}) , where the number of internal states is $M-1$ with the following pairs: $y_1 = (T_1, T_2), y_2 = (T_2, T_3), \dots, y_{M-1} = (T_{M-1}, T_M)$. We also propose the differential equation based on Eq. (11) given by

$$\frac{dy_m}{dt} = \frac{1}{1 + \exp(\Delta_m)}, \quad (12)$$

where Δ_m is given by Eq. (8).

Compared to REM, the difference of the algorithms is in Step 2. In Step 2 of DETREM, instead of evaluating the Metropolis criterion, the update of each y_i is done by the differential equation in Eq. (12). Step 2 in DETREM is given as follows:

- (i) All internal states y_m ($m = 1, \dots, M-1$) for replica exchange pairs are integrated independently based on Eq. (12). Namely, we update y_m by

$$y_m(t + \Delta t) = y_m(t) + \sigma_m \frac{1}{1 + \exp(\Delta_m)} dt, \quad (m = 1, 2, \dots, M-1) \quad (13)$$

where Δ_m is evaluated with the last coordinates in the simulations in Step 1 above and the signature σ_m of the pair of (T_m, T_{m+1}) changes to 1 or -1 to control the signature for numerical accuracy of the change of y_m which monotonically increases or decrease.

- (ii) When the value of the internal state y_m is ≥ 1 or ≤ -1 , the temperature exchange of the pair corresponding to y_m is performed as follows,

Temperature pair is exchanged :

if updated $y_m \geq 1$, then $(T_m, T_{m+1}) \rightarrow (T_{m+1}, T_m)$, $y_m \leftarrow y_m - 1$, $\sigma_m \leftarrow -1$,

if updated $y_m \leq -1$, then $(T_m, T_{m+1}) \rightarrow (T_{m+1}, T_m)$, $y_m \leftarrow y_m + 1$, $\sigma_m \leftarrow +1$.

Figure 1 summarizes the algorithm. Note that we add value 1 or -1 for new y_m after exchanges are made because of our coding. We remark that DETREM is performed just like GSREM, where in Step 2 the evaluation of the conditional probability in Eq. (11) by pseudo random numbers is replaced by solving the differential equation in Eq. (12).

The DETREM equation is proposed to yield the distributions in Eq. (11) based on the conditional probability in the heat-bath method. While spins are updated in the Chaotic Boltzmann machine[24], temperatures are updated in DETREM. The neighboring spins of the chosen spin site are assumed to be fixed in the former method, and this assumption for DETREM corresponds to the one that temperatures in unselected replicas are fixed during Step 2 above, although spins are changed during Step 1. We remark that DETREM worked properly even when we changed many spins in Step 1, while only one spin was updated in the present simulation[38].

Expectation values of physical quantities are given as functions of temperatures by WHAM[36, 37, 39]. Namely, the density of states $n(E)$ and dimensionless Helmholtz free energy are obtained by solving the following equations self-consistently:

$$n(E) = \frac{\sum_{m=1}^M N_m(E)}{\sum_{m=1}^M n_m e^{f_m - \beta_m E}}, \quad (14)$$

and

$$e^{-f_m} = \sum_E n(E) e^{-\beta_m E}, \quad (15)$$

where $N_m(E)$ and n_m are the energy histogram and the total number of samples obtained at temperature T_m , respectively. After we obtained f_m at each temperature, the expectation value of a physical quantity A at any temperature T is given by [39]

$$\langle A \rangle_T = \frac{\sum_{m=1}^M \sum_{x_m} A(x_m) \frac{1}{\sum_{l=1}^M n_l \exp(f_l - \beta_l E(x_m))} \exp(-\beta E(x_m))}{\sum_{m=1}^M \sum_{x_m} \frac{1}{\sum_{l=1}^M n_l \exp(f_l - \beta_l E(x_m))} \exp(-\beta E(x_m))}, \quad (16)$$

where x_m are the set of coordinates at temperature T_m obtained from the trajectories of the simulation.

Simulation conditions

In order to test the effectiveness of the present methods, we studied the 2-dimensional Ising model. The lattice size L in square lattice was 128. The system size N is equal to L^2 . In both methods, the update of spin states was performed by the Metropolis criterion.

For REM, replica-exchange attempt was made for every 1 MC step. 1 MC step was defined to be one Metropolis update of spins. The total number of MC steps was 100,000,000. Each data was sampled at 1,000 MC steps frequency.

To integrate Eq. (12), we used the fourth-order Runge-Kutta method, which is equivalent to Eq. (13), with virtual time step $dt = 1$. The initial value for $y_m(t = 0)$ was set to 0 for all m , while any initial value between -1 and 1 is acceptable. Because the initial values of y_m influence only the first replica exchange, the results will not depend on them as long as the total number of replica exchange is large enough. The total number of replicas was 40 and the temperatures were 1.50, 1.55, 1.60, 1.65, 1.70, 1.75, 1.80, 1.85, 1.90, 1.94, 1.98, 2.01, 2.04, 2.07, 2.10, 2.13, 2.16, 2.19, 2.22, 2.25, 2.28, 2.31, 2.34, 2.358, 2.368, 2.38, 2.40, 2.42, 2.44, 2.47, 2.51, 2.57, 2.63, 2.69, 2.75, 2.82, 2.90, 3.00, 3.10, and 3.15. Boltzmann constant k_B and the coupling constant J were set to 1. Thus, $\beta = 1/k_B T = 1/T = \beta^*$, and the (potential) energy is given by

$$E(\mathbf{s}) = - \sum_{\langle i,j \rangle} s_i s_j, \quad (17)$$

where $s_i = \pm 1$, and the summation is taken over all the nearest-neighbor pairs in the square lattice. The canonical distribution is given by

$$W(\mathbf{s}) = \frac{1}{Z} \exp(-\beta^* E(\mathbf{s})), \quad (18)$$

where Z is the partition function. In the DETREM simulation, all y_m were updated simultaneously. The multiple exchanges of temperatures at a replica were prohibited. Namely, if the neighboring internal states y_m and y_{m+1} satisfy the exchange condition ($y_m \geq 1$ or $y_m \leq -1$), only one state, e.g., y_m , was updated and only the pair (T_m, T_{m+1}) was exchanged, while y_{m+1} was not updated. For analysis, we used the R program package.[40–42]

RESULTS

Figure 2 shows the time series of y_m change in one of the temperature pairs ($m = 1$) as a function of MC steps from the DETREM simulation. We see a random walk, which results in random walks in temperature space for each replica. Figure 3(a) and Figure 3(b) show the time series of temperature change in one of the replicas (Replica 1) as a function of MC steps from the conventional REM and the DETREM simulation, respectively. They show similar behaviors with respect to random walks in temperature space. We see that all replicas take the minimum temperature many times during both simulations. Other replicas perform random walks similarly. Figures 4(a) and 4(b) show the time series of replica index at the minimum temperature of 1.5 during the REM and DETREM simulations. This shows that all replicas experienced the minimum temperature many times during the simulation. Table I lists the maximum number of tunneling events per replica, which is the number of times where the simulation visits from the lowest temperature through the highest temperature and back to the lowest temperature. These data show that the two methods have nearly the same number of tunneling counts during the simulations. All these results imply that REM and DETREM are equally efficient in sampling.

We next examine physical quantities obtained from the DETREM simulation and compare them to those from the REM simulation. Figure 5 shows the canonical energy distributions at 40 temperatures as functions of energy obtained from the REM and DETREM simulations. We see that the distributions have enough overlaps in pairs of the neighboring distributions. This ensures that the number of replicas is sufficient. The agreement between two methods implies that DETREM method produced the Boltzmann distributions at each temperature simulation just like REM did.

We next confirm the second-order phase transitions at the critical temperature of $T_c \sim 2.269$ in both methods. Figure 6(a) and Figure 6(b) show the total energy density ϵ as a function of T during the REM simulation and the DETREM simulation, respectively, where ϵ is defined by

$$\epsilon = \frac{E}{N}. \quad (19)$$

Figure 7(a) and Figure 7(b) show the specific heat C as a function of T from the REM simulation and the DETREM simulation, respectively, where C is defined by

$$C = \frac{1}{kT^2 N} (\langle E^2 \rangle - \langle E \rangle^2). \quad (20)$$

Figure 8(a) and Figure 8(b) show magnetization M as a function of T from the REM simulation and the DETREM

simulation, respectively, where M is defined by

$$M = \frac{\left| \sum_{i=1}^N s_i \right|}{N}. \quad (21)$$

Figure 9(a) and Figure 9(b) show susceptibility χ as a function of temperature during the REM simulation and the DETREM simulation, respectively, where χ is defined by

$$\chi = \frac{N}{kT} (\langle M^2 \rangle - \langle M \rangle^2). \quad (22)$$

Figure 10(a) and Figure 10(b) show the autocorrelation function of total energy density as a function of MC step lags from the REM simulation and the DETREM simulation, respectively. Here, the autocorrelation function of a physical quantity A is defined by

$$C(t) = \frac{\langle A(0)A(t) \rangle - \langle A \rangle^2}{\langle A^2 \rangle - \langle A \rangle^2} \quad (23)$$

$$= \frac{1}{n} \sum_{s=\max(1, -t)}^{\min(n-t, n)} \langle A_{s+t} - \langle A \rangle \rangle \langle A_s - \langle A \rangle \rangle, \quad (24)$$

where the average $\langle \dots \rangle$ is over the samples, t is the MC step lag, and the second line is written for discrete steps (A_t for $t = 1, 2, \dots, n$) from the $n-t$ observed pairs $(A_{1+t}, A_1), \dots, (A_n, A_{n-t})$. The decay coefficient τ (in $C(t) \propto e^{\tau t}$) is determined from a linear fit of $\log C(t)$ to lag in the region where the long-time single exponential decay is observed. This shows that the autocorrelation function of total energy density behaves similarly between REM and DETREM, and the decay coefficient τ of REM and DETREM in the Figures is -0.066 ± 0.002 and -0.080 ± 0.003 , respectively.

Figure 11(a) and Figure 11(b) show the autocorrelation function of magnetization M as a function of MC step lags from the REM simulation and the DETREM simulation, respectively. This shows that the autocorrelation of DETREM decreases faster than that of REM and that the decay coefficient τ of REM and DETREM is -0.025 ± 0.001 and -0.057 ± 0.002 , respectively. Thus, these results suggest that the autocorrelation for total energy density and magnetization in DETREM both decreases faster than that in REM.

All these physical quantities (Figs. 5 to 11) confirm that the DETREM simulation not only reproduced the results of the REM simulation in the phase transitions near the critical temperature T_c but also with faster convergence.

CONCLUSIONS

In this work, we proposed a deterministic replica-exchange method, which enables replicas to exchange their temperatures keeping their thermal equilibrium without using pseudo random numbers. We reproduced the results of REM by DETREM. This fact may be useful for parallel computing because the correlation in pseudo random number sequences in REM may cause defects in the results of MC simulations[43]. For example, some MC simulations which include association in sites may include hidden errors in their results[12]. Moreover, for molecular dynamics simulations, the use of bad seeds in stochastic thermostat caused the partial synchronization of trajectories[44]. Thus, similar synchronization of trajectories may occur by replica exchange with pseudo random numbers. Although to the best of our knowledge no work directly observing this phenomenon exists, we may encounter with this problem in the future. Moreover, DETREM can be considered to be a special application of the deep Boltzmann machine[30, 31] in machine learning. Thus, various effective method in machine learning can be introduced in order to further enhance the performance of DETREM. In addition, this new method can also be applied the transformation of conditional probability to other replica-exchange method variants. In another future work, we will introduce the multidimensional DETREM for generalized potential function including Hamiltonian replica-exchange method[45, 46].

ACKNOWLEDGMENTS

We are grateful to Drs. Yoshiharu Mori and Tetsuro Nagai for informing us the existence of Ref. [33]. Some of the computations were performed on the supercomputers at the Institute for Molecular Science, at the Supercomputer

Center, Institute for Solid State Physics, University of Tokyo, and Center for Computational Sciences, University of Tsukuba. This work was supported, in part, Grants-in-Aid for Scientific Research (A) (No. 25247071), for Scientific Research on Innovative Areas (“Dynamical Ordering & Integrated Functions”), Program for Leading Graduate Schools “Integrative Graduate Education and Research in Green Natural Sciences”, and for the Computational Materials Science Initiative, for High Performance Computing Infrastructure, and CREST “Molecular Technology for Chemical Control of Epigenetics towards Drug Discovery” from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan and Japan Science and Technology Agency (JST).

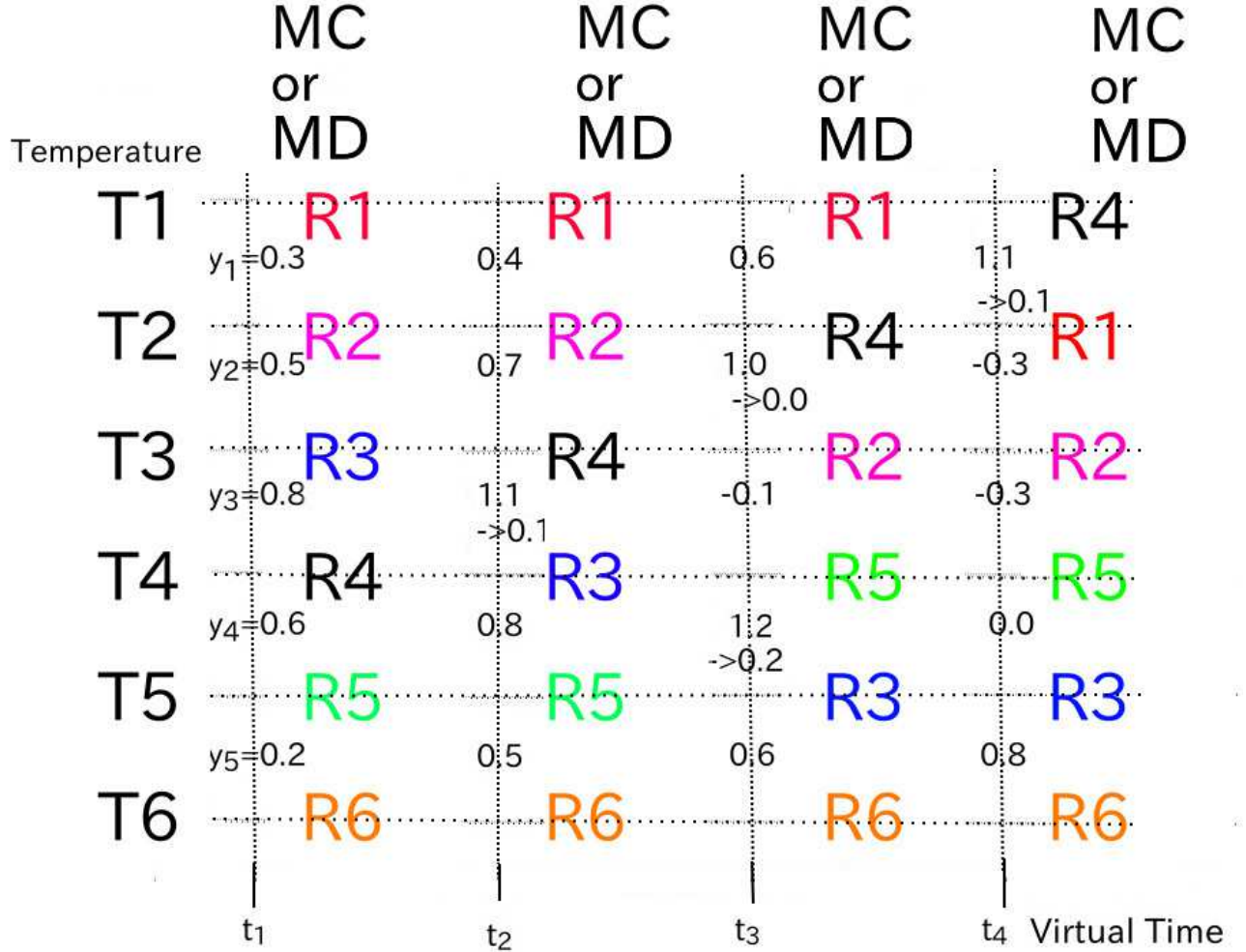


FIG. 1. Schematic sketch of DETREM. After parallel conventional MC or MD simulations at M ($=6$, here) different temperatures for short steps, all internal states y_m are updated. The temperature pairs (T_m, T_{m+1}) , or corresponding replica pairs, are exchanged when $y_m \geq 1$ or ≤ -1 . The cycle is repeated until the end of simulation. T1, T2, T3, T4, T5, and T6 are temperatures. R1, R2, R3, R4, R5, and R6 are replicas.

General Formalism of Related Methods

We present the equations for the general formalism of related methods in the following.

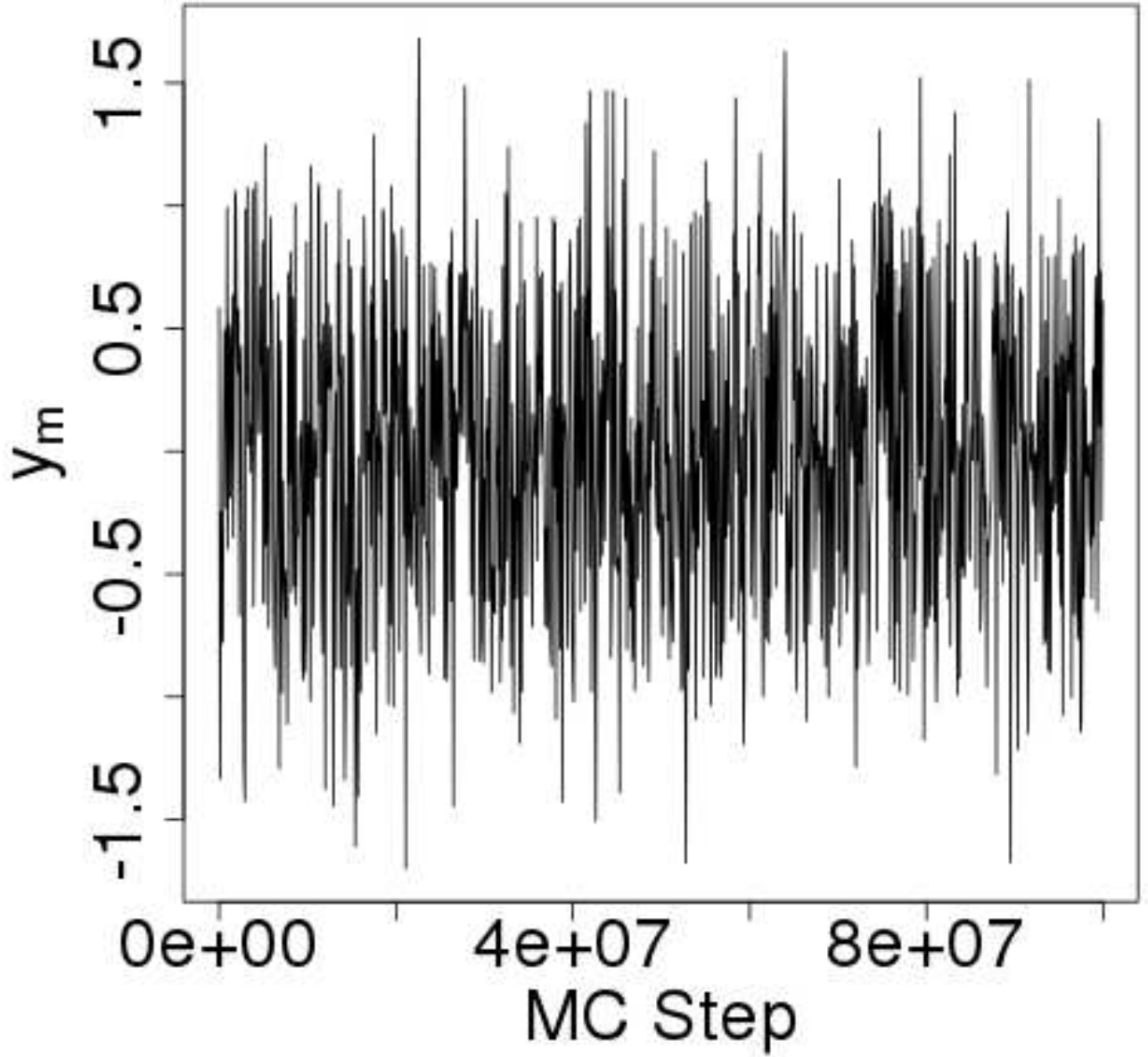


FIG. 2. Time series of y_m change in one of m ($m = 1$) for DETREM simulations. Replica exchange is performed when $y_m > 1.0$ and $y_m < -1.0$, and y_m is reset to $y_m = y_m - 1.0$ and $y_m = y_m + 1.0$, respectively,

TABLE I. The number of maximum tunneling count (TC) per replica of whole replicas during simulations

TC	REM	DETREM
Max	191	196
Mean \pm SD	173.3 ± 9.5	177.95 ± 8.8

SD means standard deviation with respect to replicas.

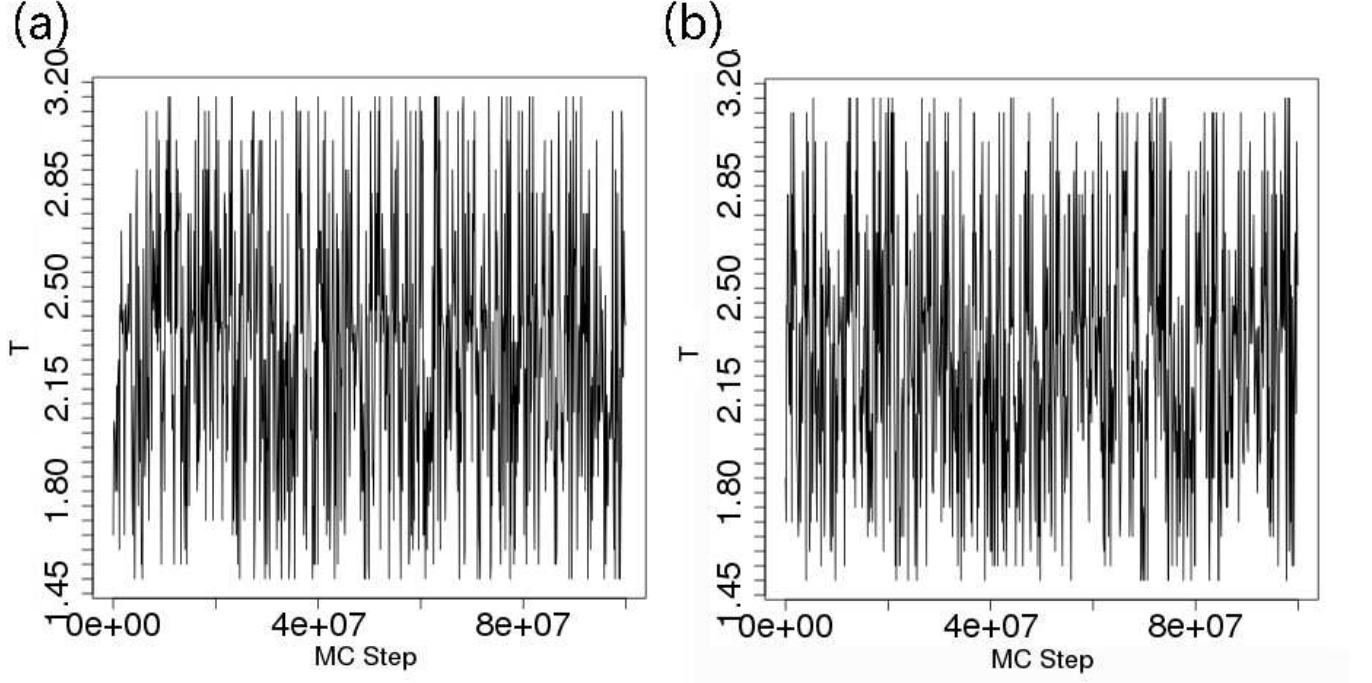


FIG. 3. Time series of temperature change in one of the replicas (Replica 1) for (a) REM and (b) DETREM simulations.

GSREM

For Gibbs sampling replica-exchange method, general equations are given in Ref. [33]. For REM with heat bath method, the conditional probability assigned for new states for any replica-exchange is given by

$$\omega(S | x^{[1]}, x^{[2]}, \dots, x^{[M]}) = \frac{W(x_{m(1)}^{[1]}, x_{m(2)}^{[2]}, \dots, x_{m(M)}^{[M]})}{\sum_{S' \in S_M} W(x_{m'(1)}^{[1]}, x_{m'(2)}^{[2]}, \dots, x_{m'(M)}^{[M]})}, \quad (25)$$

where $S \equiv \{m(1), m(2), \dots, m(M)\}$ is a permutation of temperature indices and S_M is all possible permutations.

When we consider exchange of only one pair of replicas to reduce the set S_M , Eq. (25) turns into

$$\omega(x_{m(i)}^{[i]}, x_{m(j)}^{[j]} | x_{m(k)}^{[k \neq m(i), m(j)]}) = \frac{W(x_{m(i)}^{[i]}, x_{m(j)}^{[j]})}{\sum_{i'=1}^{M-1} \sum_{j' > i'}^M W(x_{m(i')}^{[i']}, x_{m(j')}^{[j']})}. \quad (26)$$

DETREM

We propose the differential equations for DETREM corresponding to the general equations of GSREM. At first, we introduce an internal state y to assign a permutation of temperature state based on Eq. (25). It satisfies the following differential equation:

$$\frac{dy}{dt} = \frac{W(x_{m(1)}^{[1]}, x_{m(2)}^{[2]}, \dots, x_{m(M)}^{[M]})}{\sum_{S' \in S_M} W(x_{m'(1)}^{[1]}, x_{m'(2)}^{[2]}, \dots, x_{m'(M)}^{[M]})}, \quad (27)$$

where t is a virtual time, $y \in [1, N_{S_M} + 1]$, and N_{S_M} is the total number of elements in the permutation set S_M . This DETREM is performed just like GSREM, where in Step 2 the evaluation of the conditional probability in Eq. (25) is replaced by solving the differential equation in Eq. (27). DETREM gives the same probability as in Eq. (25).

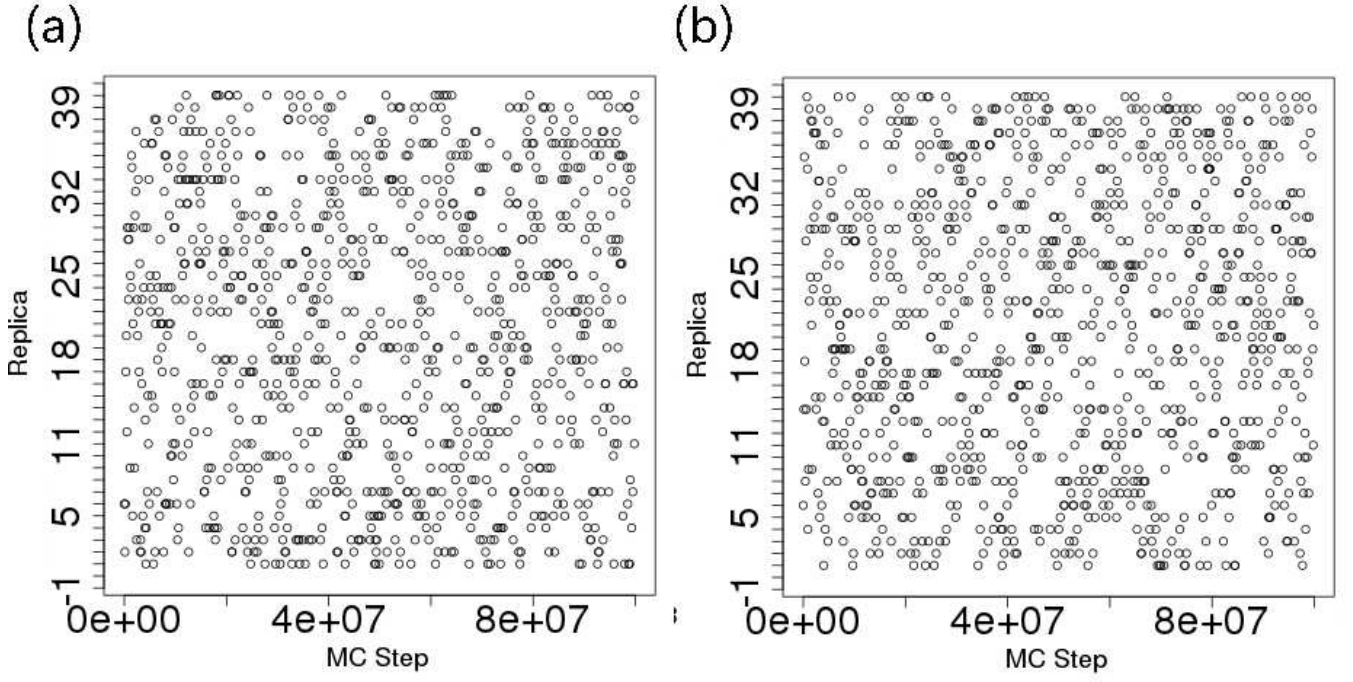


FIG. 4. Time series of replica at temperature 1.5 for (a) REM and (b) DETREM simulations.

When y reaches $m = \lfloor y \rfloor$, the permutation of temperature corresponding to the integer in temperature permutations is chosen. Here, the floor function $\lfloor y \rfloor$ is the integral part of y defined by $\lfloor y \rfloor = m \leftrightarrow m \leq y < m + 1$. In general, N_{S_M} is so large that using Eq. (27) is not practical.

We thus introduce an internal state $y_{m,n}$ for exchange of a pair of temperatures T_m and T_n that correspond to replicas i and j , respectively, with the time integration based on Eq. (26), which gives

$$\frac{dy_{m,n}}{dt} = \frac{W(x_m^{[i]}, x_n^{[j]})}{W(x_m^{[i]}, x_n^{[j]}) + W(x_n^{[i]}, x_m^{[j]})} \quad (28)$$

$$= \frac{1}{1 + \exp(\Delta_{m,n})}, \quad (29)$$

where $y_{m,n} \in [0, 1]$ and $\Delta_{m,n}$ is given by Eq. (6).

DETST

We remark that simulated tempering (ST)[49, 50] corresponding to GSREM and DETREM can also be formulated. The weight factor for ST is given by

$$W_m^{\text{ST}}(x) = \exp[-\beta_m E(x) + f_m], \quad (30)$$

where f_m are the dimensionless Helmholtz free energy at temperature T_m ($m = 1, \dots, M$). In Step 1, we perform a canonical MC or MD simulation at temperature T_m for short steps. In Step 2, when we consider temperature change to neighboring values[51], the conditional probability from temperature T_m into T_{m+1} or T_{m-1} is given by

$$\omega(T_{m\pm 1} | T_m, x) = \frac{W_{m\pm 1}^{\text{ST}}(x)}{W_{m\pm 1}^{\text{ST}}(x) + W_m^{\text{ST}}(x)} = \frac{1}{1 + \frac{W_m^{\text{ST}}(x)}{W_{m\pm 1}^{\text{ST}}(x)}} \quad (31)$$

$$= \frac{1}{1 + \Delta_{\pm}^{\text{ST}}}, \quad (32)$$

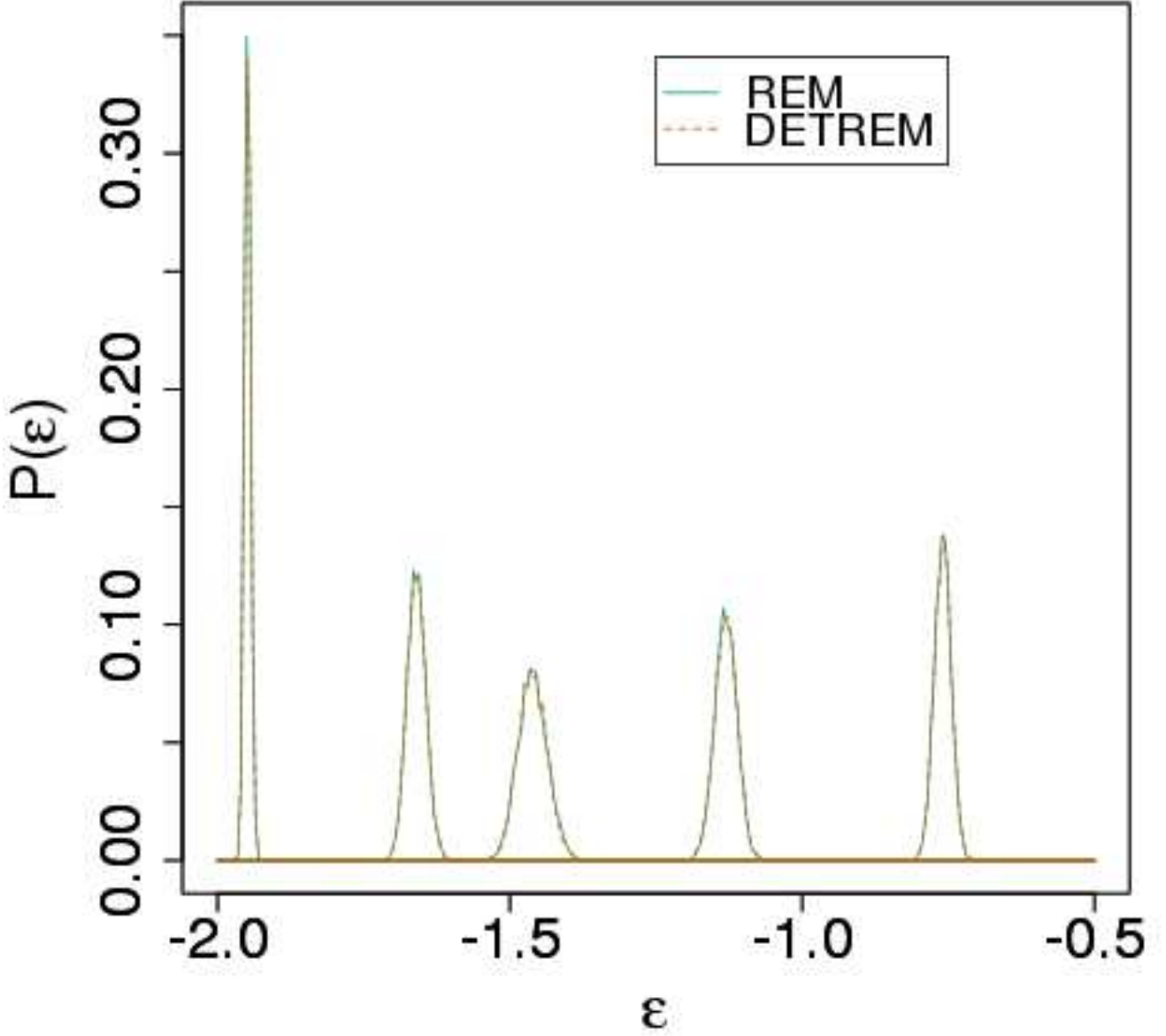


FIG. 5. Probability distributions of energy density at four temperatures (from left to right, 1.50, 2.10, 2.25, 2.47, and 3.15) obtained from (a) REM and (b) DETREM simulations including the mixed-walk simulation.

where Δ_{\pm}^{ST} are defined by

$$\Delta_{\pm}^{\text{ST}} = (\beta_{m\pm 1} - \beta_m)E(x) - (f_{m\pm 1} - f_m). \quad (33)$$

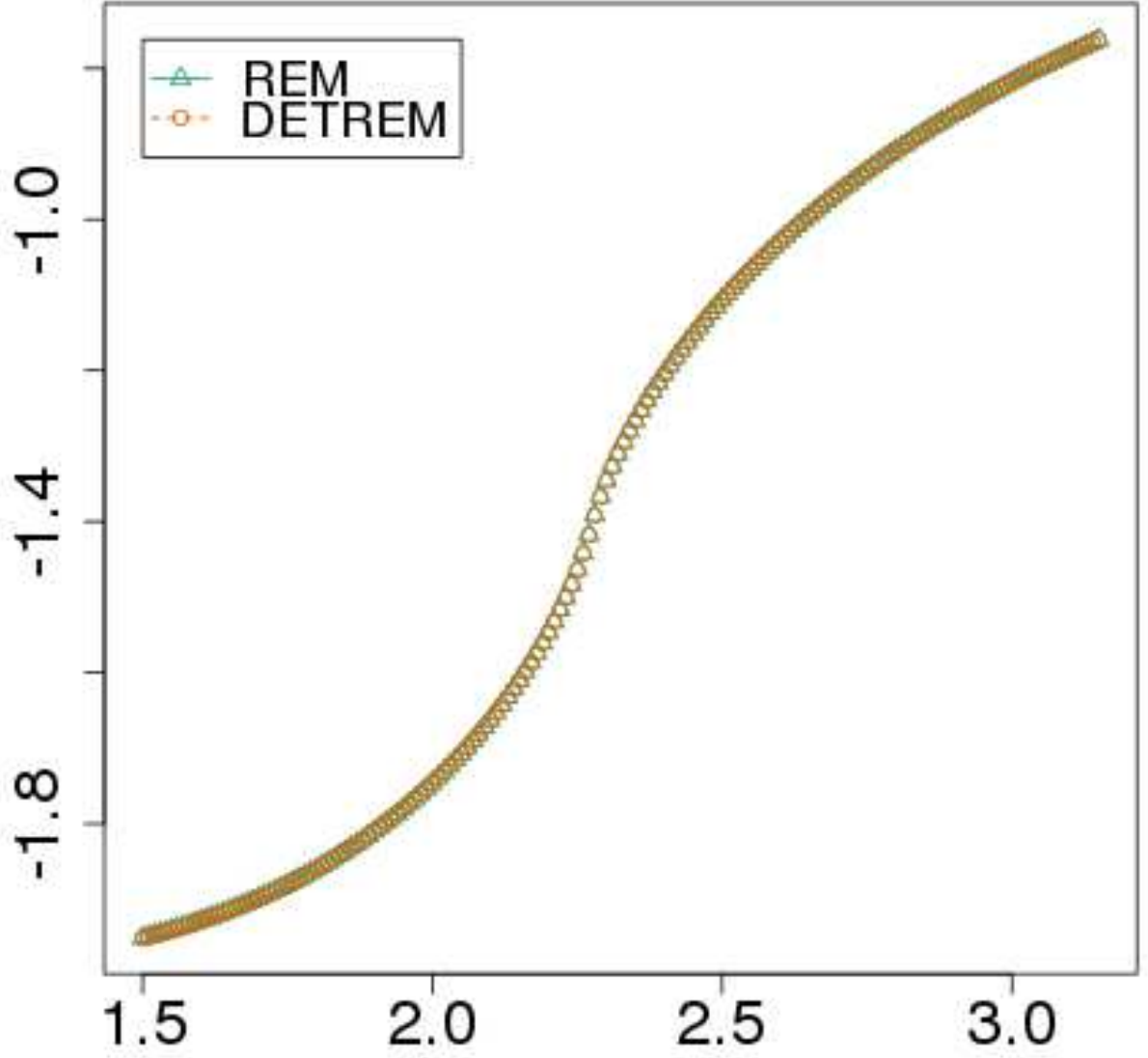


FIG. 6. Average energy density as a function of temperature obtained by WHAM from the REM and DETREM simulations. The error bars are smaller than the symbols.

This is the transition probability for the Gibbs sampling simulated tempering (GSST) (this formulation was first given in Ref. [33]). Hence, the internal states and differential equations are given by

$$\frac{dy_m^+}{dt} = \frac{1}{1 + \Delta_+^{\text{ST}}}, \quad (34)$$

$$\frac{dy_m^-}{dt} = \frac{1}{1 + \Delta_-^{\text{ST}}}, \quad (35)$$

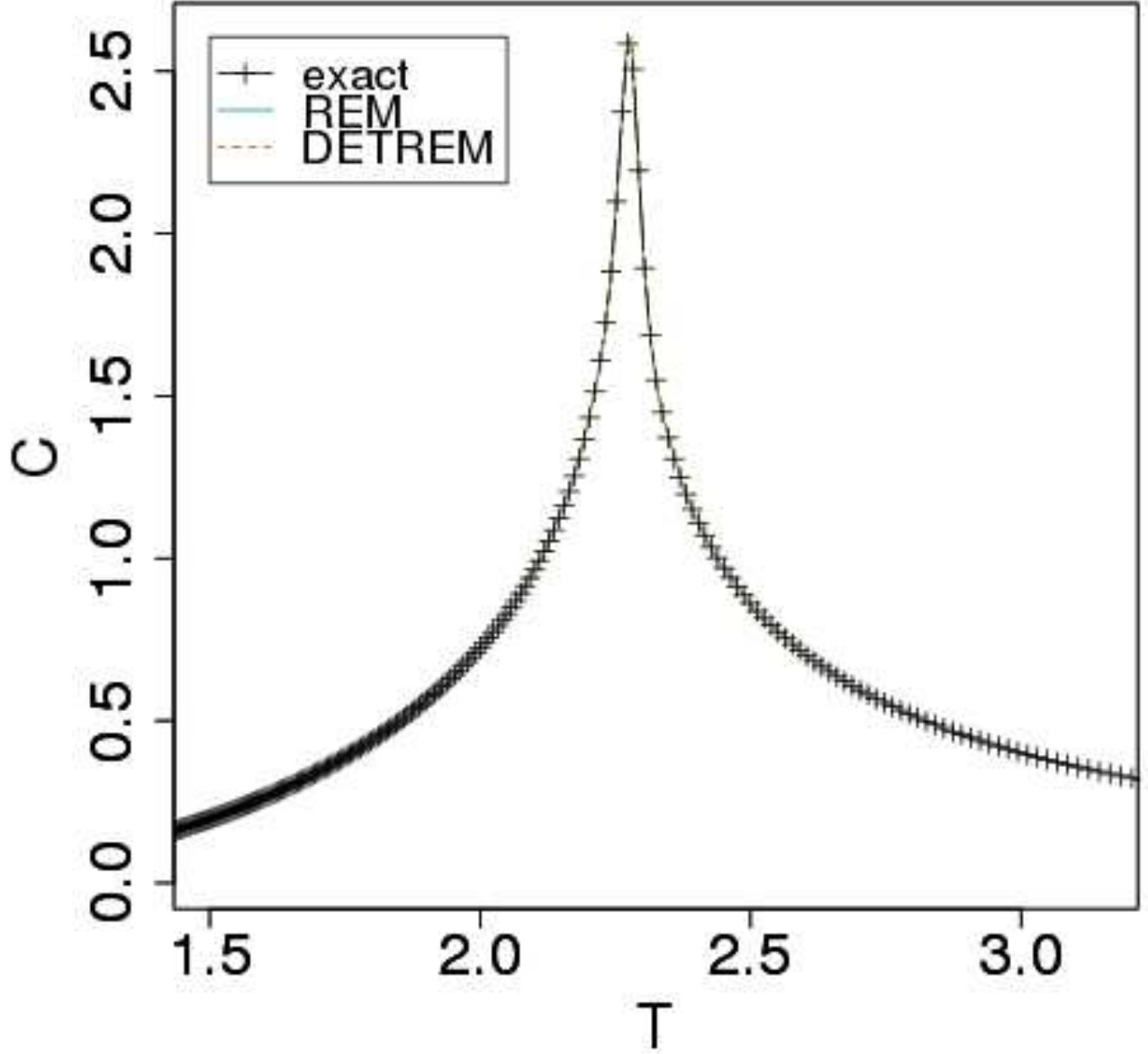


FIG. 7. Specific heat as a function of temperature obtained by WHAM from the REM and DETREM simulations. The error bars are smaller than the symbols. The exact results for $L=128$ (black curves) were obtained by Berg's program [47] based on Ref. [48].

where $y_m = y_{m,m+1}$ ($m = 1, \dots, M-1$). When the system stays at temperature T_m , only y_m^\pm are updated, and other $y_{n \neq m}^\pm$ are not updated.

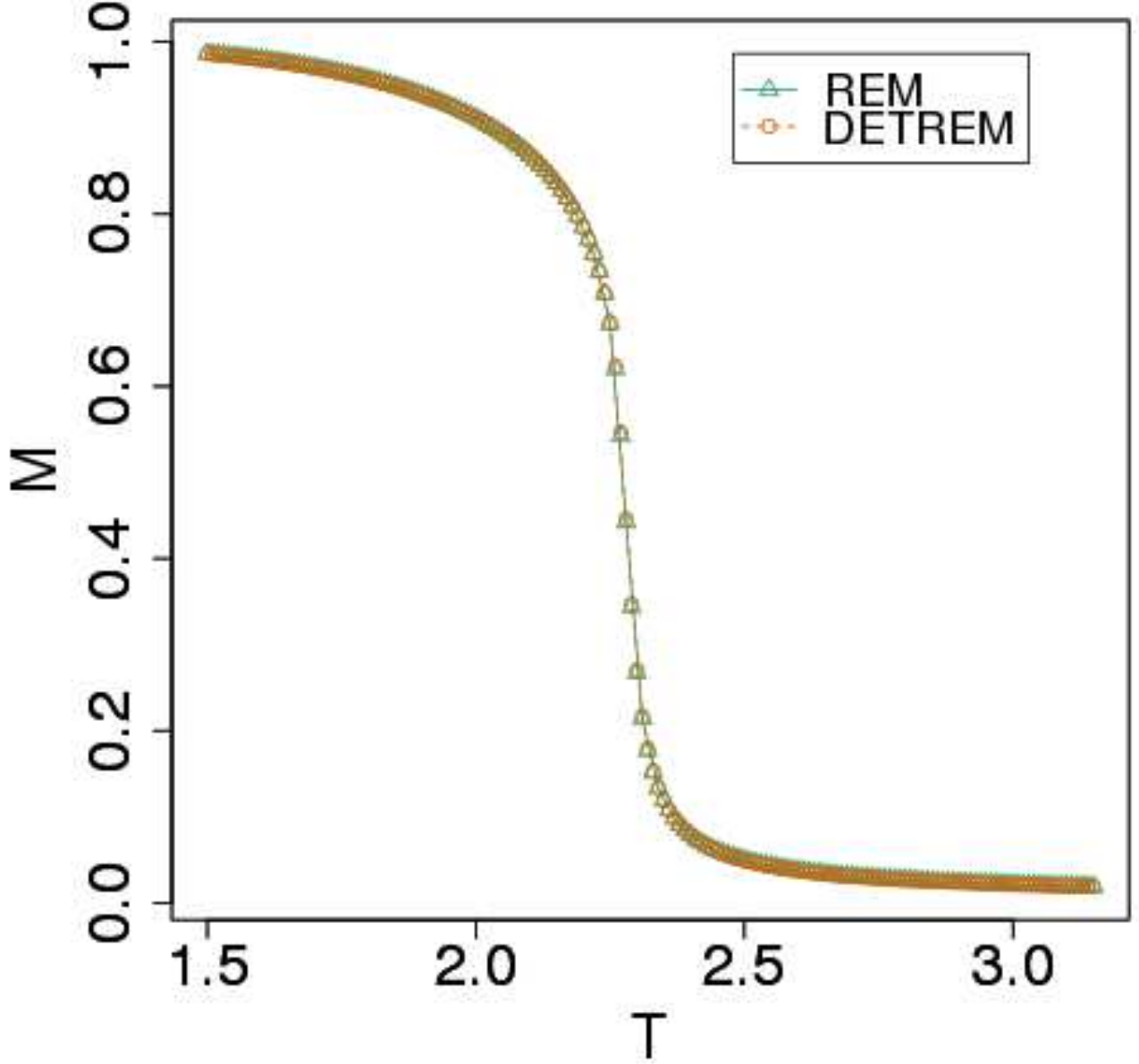


FIG. 8. Magnetization as a function of temperature obtained by WHAM from the REM and DETREM simulations. The error bars are smaller than the symbols.

New temperature is updated to a new value T_n with the following conditional probability:

$$\omega(T_n | x) = \frac{W_n^{\text{ST}}(x)}{\sum_{m=1}^M W_m^{\text{ST}}(x)} = \frac{\exp[-\beta_n E(x) + f_n]}{\sum_{m=1}^M \exp[-\beta_m E(x) + f_m]} \quad (36)$$

$$\propto \exp[-(\beta_n - \beta_0)E(x) + (f_n - f_0)], \quad (37)$$

where we have introduced an arbitrary reference temperature T_0 to give the normalization.

From above, we can derive a differential equation for deterministic simulated tempering (DETST). We introduce

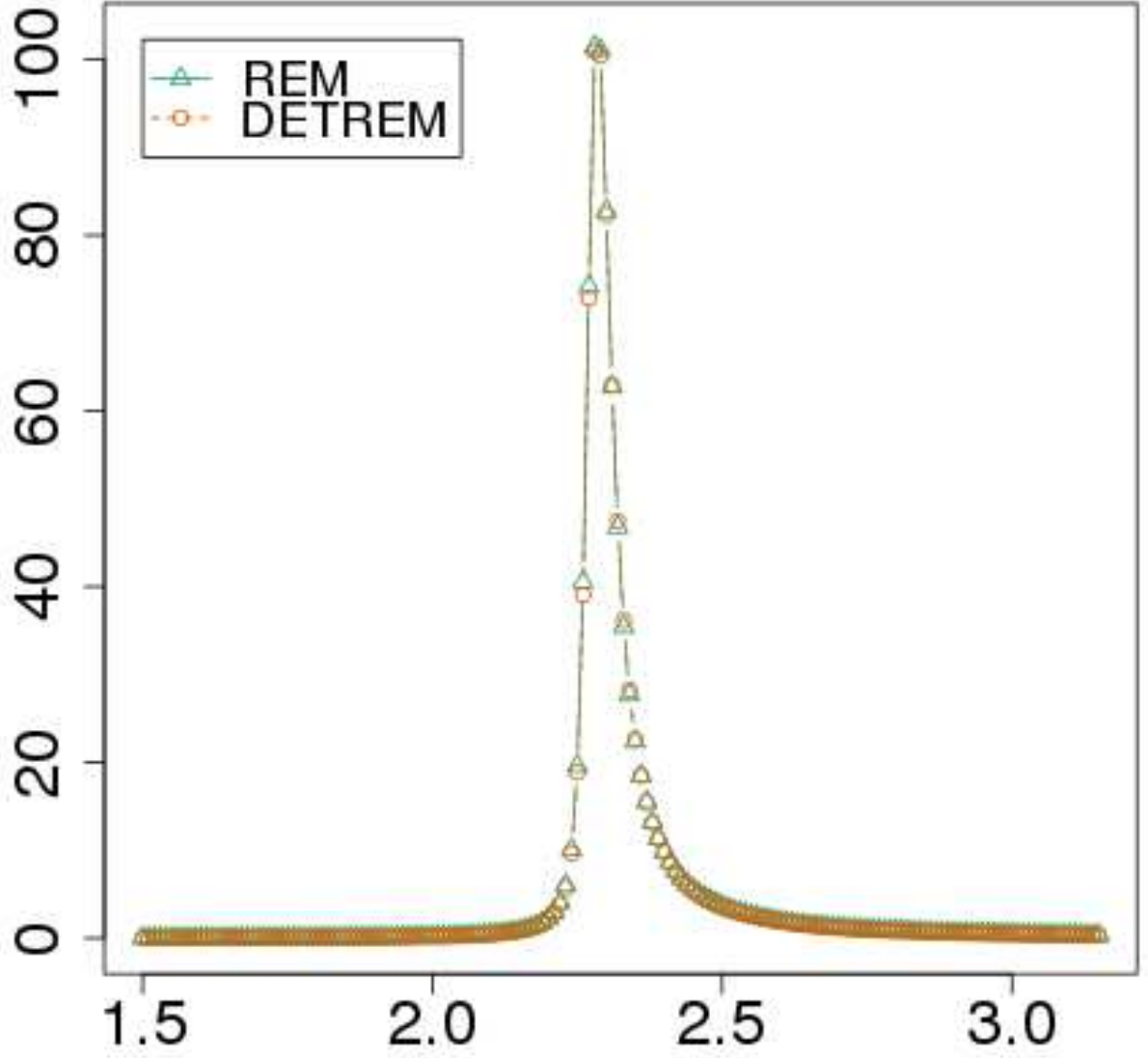


FIG. 9. Susceptibility as a function of temperature obtained by WHAM from the REM and DETREM simulations. The error bars are smaller than the symbols.

an internal state y

$$\frac{dy}{dt} = \exp [-(\beta_n - \beta_0)E(x) + (f_n - f_0)], \quad (38)$$

where $y \in \{1, \dots, M+1\}$ and $m = \lfloor y \rfloor$.

In another implementation, the conditional probability from the current temperature T_m to T_n is given by

$$\omega(T_n | x, T_m) = \frac{W_n^{\text{ST}}(x)}{W_n^{\text{ST}}(x) + W_m^{\text{ST}}(x)}. \quad (39)$$

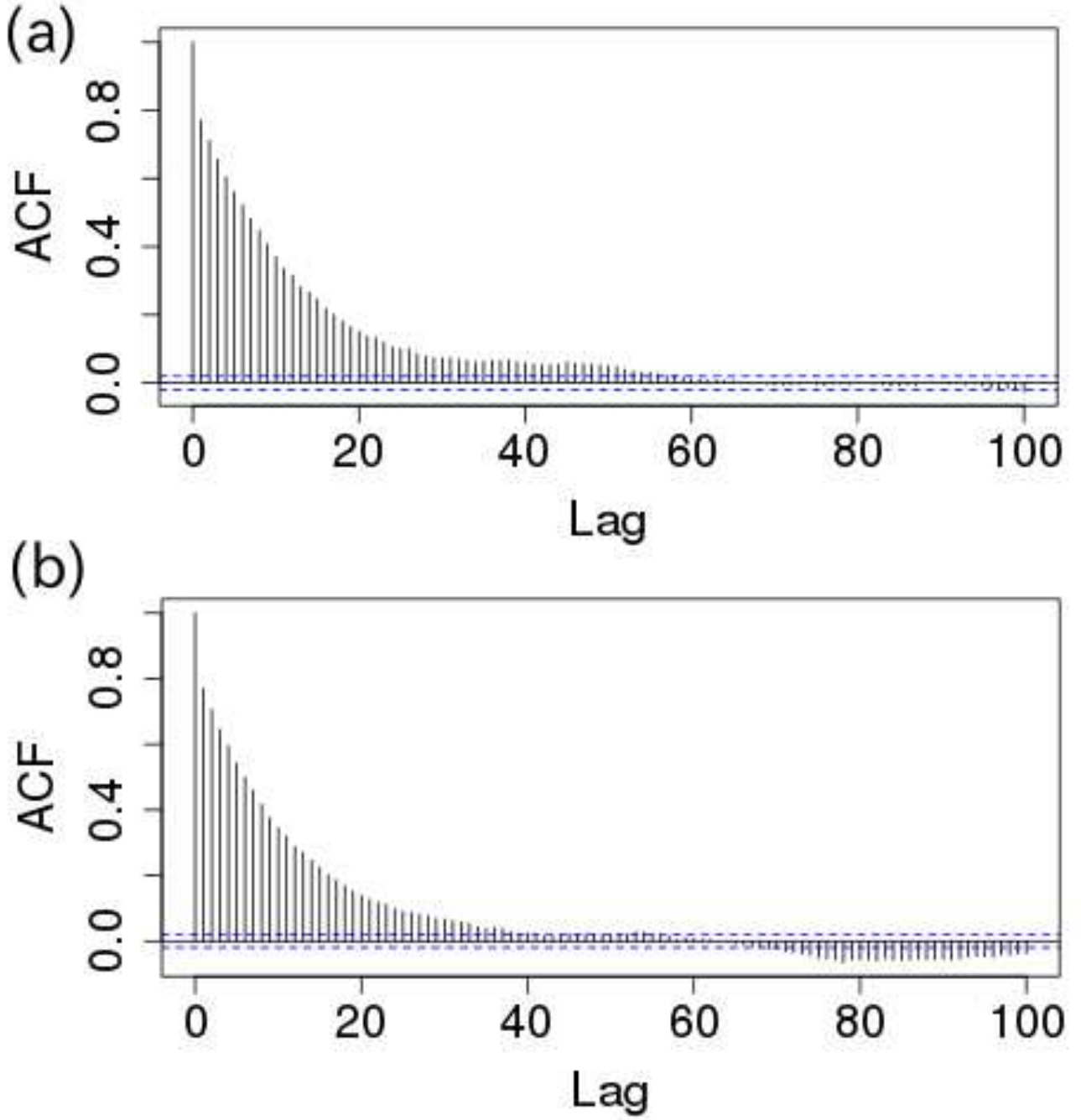


FIG. 10. Autocorrelation function of total energy density as a function of MC step lag in one of the replicas (Replica 1) for (a) REM and (b) DETREM simulations. The blue dashed lines show a 95% confidence interval.

Thus, we introduce an internal state $y_{m,n}$ integrated by

$$\frac{dy_{m,n}}{dt} = \frac{W_n^{\text{ST}}(x)}{W_n^{\text{ST}}(x) + W_m^{\text{ST}}(x)} = \frac{1}{1 + \frac{W_m^{\text{ST}}(x)}{W_n^{\text{ST}}(x)}} \quad (40)$$

$$= \frac{1}{1 + \Delta_{m,n}^{\text{ST}}} , \quad (41)$$

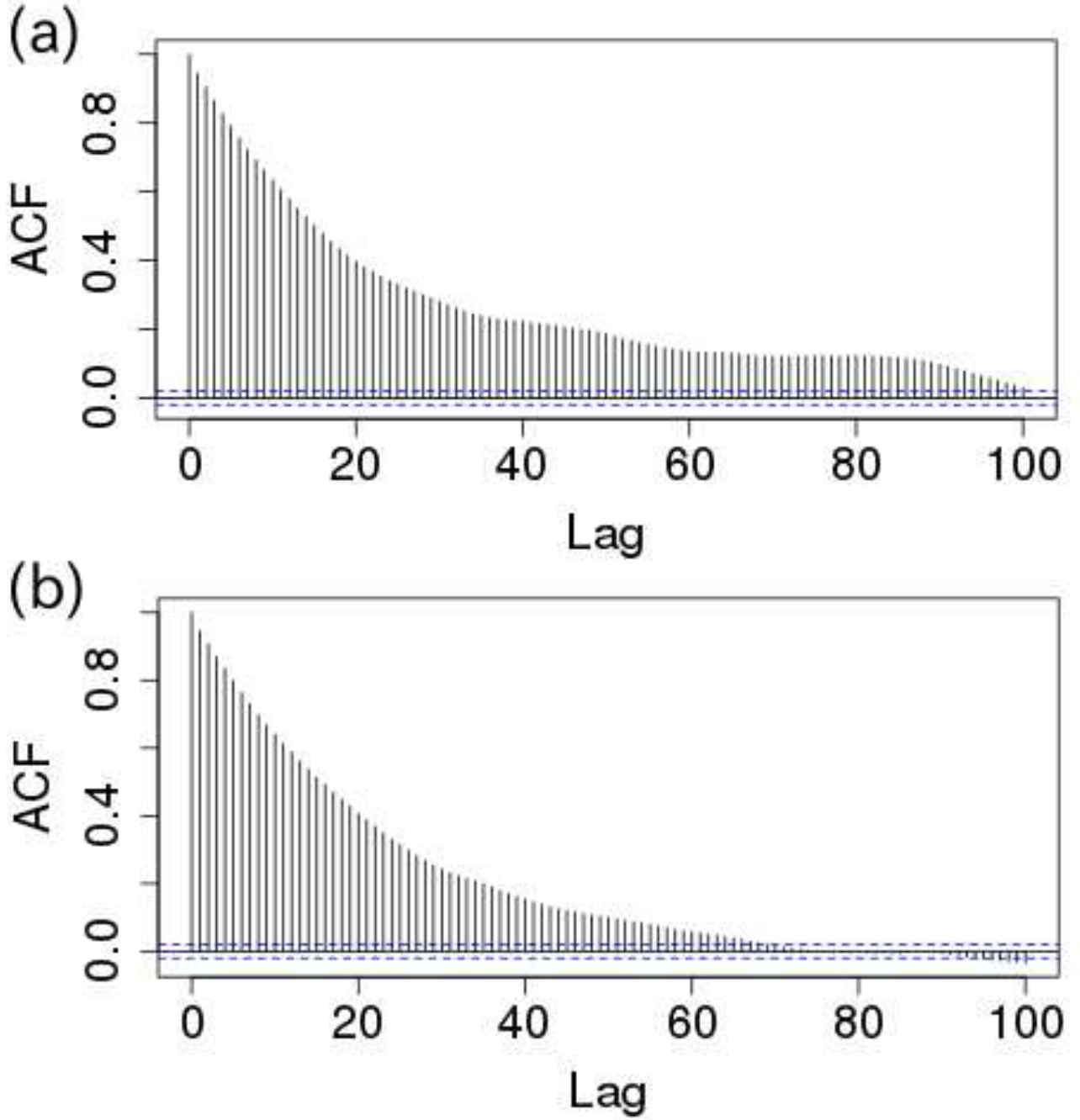


FIG. 11. Autocorrelation function of magnetization as a function of MC step lag in one of the replicas (Replica 1) for (a) REM and (b) DETREM simulations. The blue dashed lines show a 95% confidence interval.

where $y_{m,n} \in \{0, 1\}$, $m, n \in \{1, 2, \dots, M\}$, and

$$\Delta_{m,n}^{\text{ST}} = (\beta_n - \beta_m)E(x) - (f_n - f_m). \quad (42)$$

-
- [1] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
 [2] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **57**, 2607 (1986).

- [3] C. J. Geyer, *Comput. Sci. Stat.: Proc. 23rd Symp. Interface*, Interface Foundation, Fairfax Station, VA , 156 (1991).
- [4] Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999).
- [5] P. Liu, B. Kim, R. A. Friesner, and B. J. Berne, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13749 (2005).
- [6] L. Wang, R. A. Friesner, and B. Berne, *J. Phys. Chem. B* **115**, 9431 (2011).
- [7] T. Mori, J. Jung, and Y. Sugita, *J. Chem. Theory Comput.* **9**, 5629 (2013).
- [8] T. Mori, J. Jung, and Y. Sugita, *Biophys. J.* **106**, 704a (2014).
- [9] A. Mitsutake, Y. Sugita, and Y. Okamoto, *Biopolymers* **60**, 96 (2001).
- [10] Y. Iba, *Int. J. Mod. Phys. C* **12**, 623 (2001).
- [11] A. I. Georgescu, S. J. Denny, E. Joly, G. Chen, D. Perez, and A. F. Voter, *Phys. Rev. E* **86**, 026703 (2012).
- [12] A. M. Ferrenberg, D. P. Landau, and Y. J. Wong, *Phys. Rev. Lett.* **69**, 3382 (1992).
- [13] Y. Sugita, N. Miyashita, P.-C. Li, T. Yoda, and Y. Okamoto, *Curr. Phys. Chem.* **2**, 401 (2012).
- [14] T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, *Phys. Rev. Lett* **110**, 210603 (2013).
- [15] T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, *Phys. Rev. E* **90**, 023302 (2014).
- [16] T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, in *J. Phys.: Conf. Series*, Vol. 487 (IOP Publishing, 2014) p. 012001.
- [17] J. K. Salmon, M. A. Moraes, R. O. Dror, and D. E. Shaw, in *High Performance Computing, Networking, Storage and Analysis (SC), 2011 International Conference for* (IEEE, 2011) pp. 1–12.
- [18] S. Trebst, D. A. Huse, and M. Troyer, *Phys. Rev. E* **70**, 046701 (2004).
- [19] W. Nadler and U. H. E. Hansmann, *Phys. Rev. E* **76**, 065701 (2007).
- [20] H. Nymeyer, *J. Chem. Theory Comput.* **4**, 626 (2008).
- [21] M. J. Abraham and J. E. Gready, *J. Chem. Theory Comput.* **4**, 1119 (2008).
- [22] E. Rosta and G. Hummer, *J. Chem. Phys.* **131**, 165102 (2009).
- [23] H. Suzuki, J. Imura, Y. Horio, and K. Aihara, *Sci. Rep.* **3** (2013).
- [24] H. Suzuki, *Phys. Rev. E* **88**, 052144 (2013).
- [25] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, *Cognitive science* **9**, 147 (1985).
- [26] M. J. Wainwright and M. I. Jordan, *Foundations and Trends® in Machine Learning* **1**, 1 (2008).
- [27] T. Tanaka, *Phys. Rev. E* **58**, 2302 (1998).
- [28] S.-I. Amari, K. Kurata, and H. Nagaoka, *Neural Networks, IEEE Transactions on* **3**, 260 (1992).
- [29] K. Yamazaki and S. Watanabe, *Neural Networks, IEEE Transactions on* **16**, 312 (2005).
- [30] R. Salakhutdinov and G. E. Hinton, in *International Conference on Artificial Intelligence and Statistics* (2009) pp. 448–455.
- [31] R. Salakhutdinov and G. Hinton, *Neural computation* **24**, 1967 (2012).
- [32] G. Hinton, *Neural computation* **14**, 1771 (2002).
- [33] J. D. Chodera and M. R. Shirts, *J. Chem. Phys.* **135**, 194110 (2011).
- [34] H. Suwa and S. Todo, *Phys. Rev. Lett.* **105**, 120603 (2010).
- [35] S. G. Itoh and H. Okumura, *J. Chem. Theory Comput.* **9**, 570 (2013).
- [36] A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **63**, 1195 (1989).
- [37] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, *J. Comput. Chem.* **13**, 1011 (1992).
- [38] R. Urano and Y. Okamoto, e-print arXiv:1501.00772 (2015).
- [39] A. Mitsutake, Y. Sugita, and Y. Okamoto, *J. Chem. Phys.* **118**, 6664 (2003).
- [40] R Core Team, *R: A Language and Environment for Statistical Computing* (2013).
- [41] R. Ihaka and R. Gentleman, *J. Comput. Graph. Stat.* **5**, 299 (1996).
- [42] W. Venables and B. Ripley, *Modern Applied Statistics with S* (Springer-Verlag, New York, 2002).
- [43] Y. Li, M. Mascagni, and A. Gorin, *Parallel Comput.* **35**, 269 (2009).
- [44] D. J. Sindhikara, S. Kim, A. F. Voter, and A. E. Roitberg, *J. Chem. Theory Comput.* **5**, 1624 (2009).
- [45] Y. Sugita, A. Kitao, and Y. Okamoto, *J. Chem. Phys.* **113**, 6042 (2000).
- [46] H. Fukunishi, O. Watanabe, and S. Takada, *J. Chem. Phys.* **116**, 9058 (2002).
- [47] B. A. Berg, *Markov Chain Monte Carlo Simulations and Their Statistical Analysis* (World Scientific, Singapore, 2004).
- [48] A. E. Ferdinand and M. E. Fisher, *Phys. Rev.* **185**, 832 (1969).
- [49] A. Lyubartsev, A. Martsinovski, S. Shevkunov, and P. Vorontsov-Velyaminov, *J. Chem. Phys.* **96**, 1776 (1992).
- [50] E. Marinari and G. Parisi, *Europhys. Lett.* **19**, 451 (1992).
- [51] K. Shida, *BMC Bioinf.* **7**, 486 (2006).